



Density and Distribution Function estimation through iterates of fractional Bernstein Operators

C. Manté

► To cite this version:

C. Manté. Density and Distribution Function estimation through iterates of fractional Bernstein Operators. 21st International Conference on Computational Statistics, Aug 2014, Genève, Switzerland. pp.335-342. hal-01070827

HAL Id: hal-01070827

<https://hal.science/hal-01070827>

Submitted on 2 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Density and Distribution Function estimation through iterates of fractional Bernstein Operators

Claude Manté, *Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO, UM 110, Campus de Luminy Marseille, France, claudio.mante@mio.osupytheas.fr*

Abstract. We describe a method for distribution function and density estimation with Bernstein polynomials. We take advantage of results about the eigenstructure of the Bernstein operator to refine the Sevy's convergence acceleration method, based on iterates of this operator; the original Sevy's algorithm is improved by introducing fractional operators. The proposed algorithm has better convergence properties than the classical one; the price to pay is a controllable loss of the shape-preserving properties of the Bernstein approximation (monotonicity and positivity in the Density Estimation setting). The method is tested on simulated data.

Keywords. Density Estimation, Bernstein operator, root of operators, Bernstein polynomials, Lagrange polynomials

1 Introduction

Bernstein simultaneously introduced in 1912 the polynomials and the operator that bear his name in a famous paper [2]. But, as Farouki [8] noticed, this approximation has been seldom used, due to its slow convergence. For instance, to approach $f(t) = t^2$ on the unit interval with a maximal error of 10^{-4} , we need a polynomial of degree 2500 [8] ! Nevertheless, this operator (denoted B_n) has attractive shape-preserving properties: if f is positive (or monotone, or convex), its image $B_n[f]$ is so (see [5] for further properties). Consequently, the structure of a distribution function (**d.f.**) is preserved by B_n ; this point strongly motivated the use of the Bernstein approximation in Density Estimation [19, 1, 3, 12, 13, 14].

Notations

We will work in the Banach space $C[0, 1]$ of continuous functions on $[0, 1]$, equipped with the norm $\|f\| := \max_{x \in [0, 1]} |f(x)|$. The subspace of polynomials of degree $\leq n$ will be denoted \mathfrak{P}_n , and

$\overline{\mathfrak{P}_n}$ will be the supplementary of \mathfrak{P}_1 in \mathfrak{P}_n . We will denote \mathcal{F}^+ the closed convex cone of positive functions of $C[0, 1]$, and \mathcal{F}^1 the closed convex set of functions of $C[0, 1]$ integrating to 1.

Consider an operator $U : C[0, 1] \rightarrow C[0, 1]$; for $n \geq 2$ (fixed), its restriction to \mathfrak{P}_n will be denoted $\overset{\circ}{U}$, and its restriction to $\overline{\mathfrak{P}_n}$ will be denoted \overline{U} . For the sake of simplicity, the restrictions of the identity operator to these subspaces will be denoted 1 instead of $\overset{\circ}{1}$ or $\overline{1}$; $Mat(U; B_1, B_2)$ will denote the matrix representation of U with respect to the bases B_1 and B_2 of $C[0, 1]$. We will use the matrix p -norms $\|U\|_p := \sup_{v \neq 0} \frac{\|U(v)\|_p}{\|v\|_p}$ where $\|v\|_p$ is the usual vector ℓ^p -norm. Notice that $\|U\|_1 := \max_{1 \leq k \leq \dim(U)} \sum_{j=1}^{\dim(U)} |U_{j,k}|$, $\|U\|_2$ is the spectral norm, and $\|U\|_\infty := \max_{1 \leq j \leq \dim(U)} \sum_{k=1}^{\dim(U)} |U_{j,k}|$ (see [7]).

2 Expression of powers of the Bernstein operator into different bases

The Bernstein operator $B_n : C[0, 1] \rightarrow C[0, 1]$ is defined by [4, 15, 18]:

$$B_n[f](x) := \sum_{j=0}^n w_{n,j}(x) f\left(\frac{j}{n}\right)$$

with $w_{n,j}(x) := \binom{n}{j} x^j (1-x)^{n-j}$; its range $\mathcal{R}(B_n) \subseteq \mathfrak{P}_n$. Cooper and Waldron [4] gave its spectral decomposition, which can be also written into the form hereunder [16].

Theorem 2.1. *The Bernstein operator can be represented in the diagonal form*

$$B_n[f] = \sum_{j=0}^n \lambda_j^{[n]} \pi_j^{[n]} \otimes \pi_j^{*[n]} (\mathcal{L}_n[f])$$

where $f \in C[0, 1]$, $\lambda_j^{[n]} = \frac{n!}{(n-j)! n^j} \in]0, 1]$ and $\pi_j^{[n]} \in \mathfrak{P}_n$ are its eigenvalues and eigenvectors, $\pi_j^{*[n]}$ is the dual vector of $\pi_j^{[n]}$, and $u \otimes v^*(w) := u \langle v^*, w \rangle$.

We will need the Lagrange interpolation operator (**equispaced case**) $\mathcal{L}_n : C[0, 1] \rightarrow C[0, 1]$ defined by: $\mathcal{L}_n[f](x) := \sum_{j=0}^n \ell_{n,j}(x) f\left(\frac{j}{n}\right)$, with

$$\ell_{n,j}(x) := \prod_{\substack{k=0 \\ k \neq j}}^n \frac{n x - k}{j - k}.$$

Three bases of \mathfrak{P}_n will be needed:

1. the Bernstein's basis $W_n := \{w_{n,j}(x), 0 \leq j \leq n\}$
2. the Lagrange's basis $L_n := \{\ell_{n,j}(x), 0 \leq j \leq n\}$
3. the eigenvectors of B_n , $\Pi_{[n]} := \{\pi_j^{[n]}(x), 0 \leq j \leq n\}$.

Let us denote $LW_{[n]}$ the transformation matrix associated with the bases L_n and W_n , whose j^{th} column consists in the coordinates of $w_{n,j}$ in the basis L_n . The following results can be easily demonstrated [16]:

Lemma 2.2. $Mat \left(\overset{\circ}{B}_n; L_n, W_n \right) = I_n$ and $Mat \left(\overset{\circ}{B}_n; W_n, W_n \right) = LW_{[n]}$.

Thank to this lemma, we obtain for any $k \geq 2$ a first matrix representation of B_n^k from the diagram:

$$B_n^k : C[0, 1] \xrightarrow{\mathcal{L}_n} (\mathfrak{P}_n, L_n) \xrightarrow{I_n} (\mathfrak{P}_n, W_n) \xrightarrow{LW_{[n]}^{k-1}} (\mathfrak{P}_n, W_n). \quad (1)$$

Besides, Theorem 2.1 gives an alternative representation of this operator:

$$B_n^k : C[0, 1] \xrightarrow{\mathcal{L}_n} (\mathfrak{P}_n, L_n) \xrightarrow{L\Pi_{[n]}} (\mathfrak{P}_n, \Pi_{[n]}) \xrightarrow{\Lambda_{[n]}^k} (\mathfrak{P}_n, \Pi_{[n]}) \xrightarrow{\Pi W_{[n]}} (\mathfrak{P}_n, W_n) \quad (2)$$

where $\Lambda_{[n]}$ is the diagonal matrix associated with the vector

$$(1, 1, 1 - 1/n, (3n - 2)/n^2, \dots, n!/n^n)$$

of eigenvalues of B_n , and $L\Pi_{[n]}$ and $\Pi W_{[n]}$ are transformation matrices associated with the three bases.

3 Sevy's sequences for d.f. and density approximation

We saw that in the elementary case $f(t) = t^2$, the speed of convergence of $B_n[f]$ towards f is only $O\left[\frac{1}{n}\right]$ [8]; the situation is worse in the special case of d.f.s, since it can be proven [15] that one should rather expect $O\left[\frac{1}{\sqrt{n}}\right]$. To get a sequence of approximations converging faster than B_n , Sevy [17] proposed to supersede B_n by the iterated operator

$$\mathfrak{J}_n^I := \left(1 - (1 - B_n)^I\right) \quad (3)$$

and proved the following result.

Theorem 3.1. ([18], see also [4]) For $n \geq 1$ fixed, and any function F defined on $[0, 1]$, we have:

$$\|\mathfrak{J}_n^I[F] - \mathcal{L}_n[F]\| \xrightarrow{I \rightarrow \infty} 0.$$

Such a sequence build a bridge between $\mathfrak{J}_n^1[F] = B_n[F]$ and $\mathcal{L}_n[F]$. It is worth noting that $\mathcal{L}_n[F]$ interpolates the data but can be very bumpy and that in the equispaced case, the interpolation errors are maximal ([6, Ch. 2]; [11, Ch. 5]). Suppose now F is a d.f.; $B_n[F]$ is also a d.f., but in general $\mathcal{L}_n[F]$ will not share the same characteristics. Thus, it is natural to try to determine some optimal number of iterations $I^* \geq 1$ in order that $\mathfrak{J}_n^{I^*}[F]$ has the structure of a d.f., while $\mathfrak{J}_n^{I^*+1}[F]$ has not. In other words, the density approximation $\widehat{f_n^{(I^*)}}(x) := \frac{d}{dx} \mathfrak{J}_n^{I^*}[F](x)$ should be *bona fide*, i.e. should belong to $\mathcal{F}^+ \cap \mathcal{F}^1$, while $\widehat{f_n^{(I^*+1)}} \notin \mathcal{F}^+ \cap \mathcal{F}^1$ (see [15]).

4 Interpolating Sevy's sequences (see [16])

To refine Sevy's sequences, we build for $K \geq 2$ the K^{th} "root" of the operator $G_n := (1 - B_n)$ involved in Formula 3. Because B_n only preserves \mathfrak{P}_1 , the eigenvalues of $\overline{G_n}$ belong to $]0, 1[$. Thus, thanks to classical results about convergent series of operators (see [10] for instance), one may consistently define the fractional operator

$$\overline{G_n}^{(1/K)} := \exp \left(\frac{1}{K} \log (\overline{G_n}) \right). \quad (4)$$

One can easily verify the following lemma.

Lemma 4.1. $\forall I \geq 1$,

$$\mathfrak{I}_n^I = \left(1 - (1 - B_n)^I \right) = \left(1 - \left(1 - \overset{\circ}{B}_n \right)^I \right) \circ \mathcal{L}_n.$$

Consequently, we can proceed as if $f \in \mathcal{R}(\mathcal{L}_n)$ and don't have to worry about the "Lagrange residual" $f - \mathcal{L}_n[f]$. Since \mathfrak{P}_1 is preserved by B_n and because of Lemma 4.1, $\mathfrak{I}_n^k(f) = \mathcal{L}_1[f] + \mathfrak{I}_n^k(\mathcal{L}_n[f] - \mathcal{L}_1[f])$, and we can set the definition of K-fractional Sevy's sequences.

Definition 4.2. Let $K \geq 2$, and $f \in C[0, 1]$. The K -fractional Sevy's sequence of approximations of f is:

$$\mathfrak{I}_{n;K}^j[f] := \mathcal{L}_1[f] + \left(1 - \overline{G_n}^{(j/K)} \right) (\mathcal{L}_n[f] - \mathcal{L}_1[f]), \quad j \geq 1.$$

Such a sequence interpolates the original one, since $\mathfrak{I}_{n;K}^{jK}[f] = \mathfrak{I}_n^j(f)$. Its matrix representation stems from diagram 2.

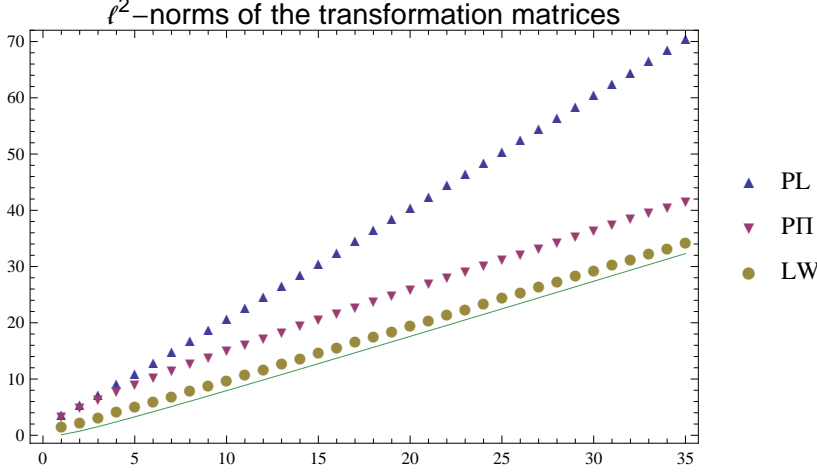
Lemma 4.3. Mat $\left(\overset{\circ}{\mathfrak{I}}_{n;K}^j; L_n, W_n \right) = \Pi W_{[n]} \circ \Lambda_{[n]}^{(j/K)} \circ L \Pi_{[n]}$, where $\Lambda_{[n]}^{(j/K)}$ is the diagonal matrix associated with the vector

$$\left(1, 1, 1 - \left(\frac{1}{n} \right)^{(j/K)}, 1 - \left(\frac{3n-2}{n^2} \right)^{(j/K)}, \dots, 1 - \left(1 - \frac{n!}{n^n} \right)^{(j/K)} \right).$$

5 Numerical issues

Because of Lemmas 2.2 and 4.1, building a classical Sevy's sequence amounts to compute powers of the transformation matrix $LW_{[n]}$ (see diagram 1). The condition number of this matrix in the ℓ^2 -norm is [7]: $\frac{\|LW_{[n]}\|_2}{\|LW_{[n]}^{-1}\|_2} = \frac{\lambda_0^{[n]}}{\lambda_n^{[n]}} = \frac{n^n}{n!} \approx \frac{e^n}{\sqrt{2\pi n}}$ (asymptotically - see [9]). Thus, $LW_{[n]}$ is ill-conditioned, and one must expect to meet numerical problems when n is big. The situation is potentially worse for fractional sequences, since Lemma 4.3 shows that the matrix of the restricted operator depends on both the ill-conditioned transformation matrices $L \Pi_{[n]}$ and $\Pi W_{[n]}$ (see Figure 1).

Figure 1. Logarithm of the condition numbers of the transformation matrices $PL_{[n]}$, $P\Pi_{[n]}$ and $LW_{[n]}$; the continuous line corresponds to the asymptotic value $n - \frac{1}{2}\text{Log}(2\pi n)$.



But the point for us is merely to control numerical errors in computing $\mathfrak{I}_{n;K}^j[f]$! Notice that on the one hand $\text{Mat}\left(\overset{\circ}{B}_n; L_n, W_n\right) = I_n$ (Lemma 2.2), while on the other hand $\text{Mat}\left(\overset{\circ}{B}_n; L_n, W_n\right) = \Pi W_{[n]} \circ \Lambda_{[n]} \circ L\Pi_{[n]}$ (diagram 2). Consequently, the matrix norms

$$\begin{aligned} & \left\| \Pi W_{[n]} \circ \Lambda_{[n]} \circ L\Pi_{[n]} - I_n \right\|_1 \\ & \left\| \Pi W_{[n]} \circ \Lambda_{[n]} \circ L\Pi_{[n]} - I_n \right\|_\infty \end{aligned} \quad (5)$$

are convenient indicators of loss of numerical accuracy imputable to the ill-conditioning of the transformation matrices. Since the only easy-to-handle basis is the power basis, the transformation matrices $PL_{[n]}$, $P\Pi_{[n]}$ and $PW_{[n]}$ are straightforwardly computed, and we can write:

$$\begin{aligned} \Pi W_{[n]} &= P\Pi_{[n]}^{-1} \circ PW_{[n]} \\ L\Pi_{[n]} &= PL_{[n]}^{-1} \circ P\Pi_{[n]} \end{aligned} \quad (6)$$

(formally). But we can derive from Figure 1 that these inverse matrices cannot be computed with sufficient accuracy in general. Thus, it's necessary to supersede in (6) the inverse matrices by the Moore-Penrose generalized inverses $P\Pi_{[n]}^+$ and $PL_{[n]}^+$. This gives rise to the regularized operators:

$$\begin{aligned} \widetilde{\Pi W}_{[n]} &:= P\Pi_{[n]}^+ \circ PW_{[n]} \\ \widetilde{L\Pi}_{[n]} &:= PL_{[n]}^+ \circ P\Pi_{[n]}. \end{aligned} \quad (7)$$

On Figure 2, we plotted the logarithm of the second indicator of Formula (5), for n ranging from 1 to 35 (a similar graph can be obtained for the first indicator). Two cases must be distinguished on this plot: the “symbolic” one, where polynomial eigenfunctions were computed from the recurrence formula given by [4], and the “numerical” one, where they were computed by polynomial interpolation of the eigenvectors of $LW_{[n]}$, giving rise to the alternative basis $\widehat{\Pi}_{[n]} := \left\{ \hat{\pi}_j^{[n]}(x), 0 \leq j \leq n \right\}$. Of course, we should have $\hat{\pi}_j^{[n]} = \pm \pi_j^{[n]} \forall 0 \leq j \leq n$ if there were

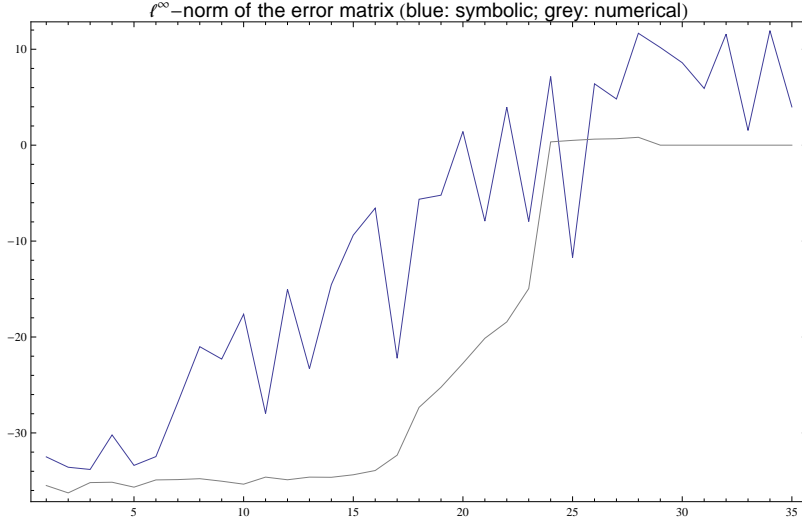


Figure 2. The sequences $\left\{ \text{Log} \left(\left\| \widetilde{\Pi} \widetilde{W}_{[n]} \circ \Lambda_{[n]}^{Tr} \circ \widetilde{L} \widetilde{\Pi}_{[n]} - I_n \right\|_{\infty} \right), 1 \leq n \leq 35 \right\}$ and $\left\{ \text{Log} \left(\left\| \widetilde{\Pi} \widetilde{W}_{[n]} \circ \Lambda_{[n]}^{Tr} \circ \widetilde{L} \widetilde{\Pi}_{[n]} - I_n \right\|_{\infty} \right), 1 \leq n \leq 35 \right\}$; $\Lambda_{[n]}^{Tr}$ is the diagonal matrix obtained by setting to zero each eigenvalue $< 10^{-12}$.

not different roundoff errors on both sides, imputable to different algorithms! That is why we took into account the numerical rank of \widetilde{B}_n , discarding from the computation of Formula (5) eigenvectors associated with eigenvalues smaller than 10^{-12} (see Figure 2 and its legend).

It is worth noting that the computational cost in the symbolic case is considerable: it took about 6600 seconds to produce the symbolic part of Figure 2, while the numeric part was obtained in 80 seconds.

6 Application to density and d.f. estimation

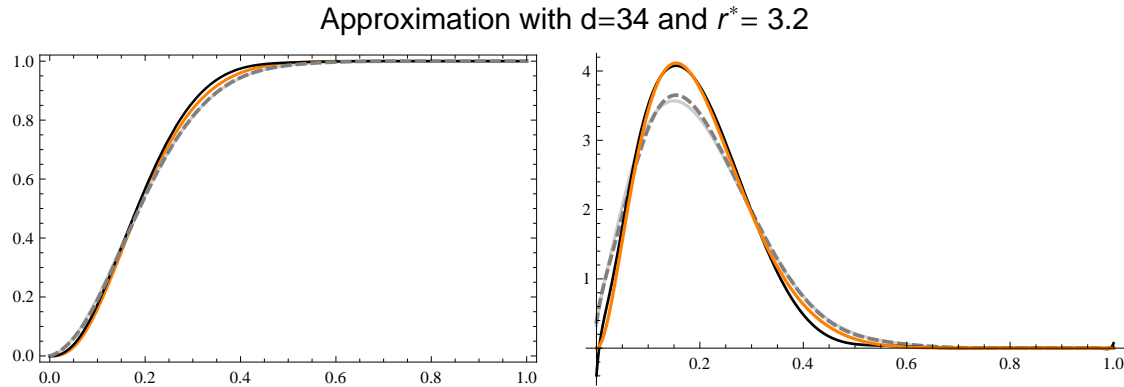
Suppose F is some differentiable d.f. associated with a random variable X defined on $[0, 1]$, and that $S_N := \{X_1, \dots, X_N\}$ is a N -sample of X , giving rise to the empirical d.f. $F_N(x)$. Babu *et al.* [1] proposed to estimate F by a Bernstein polynomial $\widetilde{F}_{N,m}$ of degree m :

$$\widetilde{F}_{N,m}(x) := \sum_{k=0}^m F_N\left(\frac{k}{m}\right) w_{m,k}(x) = B_m[F_N] \quad (8)$$

with $m \leq m_0 := \lceil N/\text{Log}(N) \rceil$. The proposed method consists in superseding $B_{m_0}[F_N]$ by some $\mathcal{J}_{m^*,K}^{I^*}[F_N]$, where $m^* \leq m_0$ and $I^* \geq K$ (fixed) are convenient values of the degree of the estimator and of the number of iterations in Definition 4.2.

As an illustration, we displayed first on Figure 3 the results obtained with a sample of size 200 of $\beta(3, 12)$, with $K = 10$. We found that $I^* = 32$ iterations of the fractional operator (4) simultaneously corresponded to a satisfactory fit of the e.d.f. and an approximately *bona fide* density estimation. Thus, in this case, the fractional number of iterations was $r^* = 1 + \frac{22}{10}$. On this plot, we superimposed to the true d.f. three estimators: the Babu's one, of degree $m_0 = 38$,

Figure 3. Estimation of the $\beta(3, 12)$ d.f. and density from a sample. Left panel: the true d.f. (orange), the Babu's one (gray and dashed, of degree $m_0 = \lceil 200/\log(200) \rceil = 38$), the classical Bernstein estimator of degree $m = 34$ (gray), and the proposed one (black), of degree 34 too. Right panel: density estimators obtained by deriving the d.f.s estimated.



the Bernstein estimator of degree $m = 34$, and the iterated estimator (black), of degree 34 too. The density estimators are derivatives of these d.f.s

In addition, we collected in Table 1 results from simulations carried on with 30 samples of size $N = 150$ ($\Rightarrow m_0 = 30$) of four Beta distributions. For sake of simplicity, we fixed $I^* = 20$ (see [16] for a theoretical justification). For each one of these samples and for each estimator (4 estimators of the d.f. and 3 estimators of the density, since the e.d.f. is not differentiable), the Integrated Squared Error (ISE) $\int (\hat{F}(x) - F(x))^2 dx$ and the L^1 error norm $\int |\hat{f}(x) - f(x)| dx$ were computed. Clearly, even in this suboptimal situation ($I^* = 20$), the proposed estimators outperformed classical ones, excepted in the very simple case $\beta(1, 2)$ (uniform distribution). Notice the honorable performances of the good old e.d.f.! ¹

Table 1. Simulations results. First group of columns: the distribution simulated, and optimal value of m (for further details, see [16]); second group: median of $10^3 \cdot ISE$ of estimated distribution functions; third group: median of the L^1 error norms for estimated densities. Best result are in bold characters.

<i>Probability</i>	m^*	<i>e.d.f.</i>	B_{30}	B_{m^*}	$\mathfrak{I}_{m^*}^{20}$	B'_{30}	B'_{m^*}	$\mathfrak{I}'_{m^*}^{20}$
$\beta(1, 2)$	16	0.497	0.415	0.38	0.569	0.1	0.09	0.108
$\beta(2, 4)$	18	0.6	0.51	0.56	0.368	0.108	0.12	0.099
$\beta(3, 12)$	25	0.32	0.783	0.908	0.258	0.197	0.207	0.118
$\beta(10, 10)$	25	0.318	1.16	1.37	0.289	0.248	0.263	0.153

Bibliography

- [1] Babu, G. J., Canty, A. J. and Chaubey, Y. P. (2002) *Application of Bernstein polynomials for smooth estimation of a distribution and density function*. Journal of Statistical Planning

¹In Section 6, the default iteration number has been erroneously set to 20. The correct value is indeed 2!

- and Inference, **105**, 377-392.
- [2] Bernstein, S. N. (1912) *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*. Commun. Soc. Math. Kharkov, **13**, 1-2.
 - [3] Bouezmarni, T. and Rolin, J.M. (2007) *Bernstein estimator for unbounded density function*. Journal of Nonparametric Statistics, **19**, **3**, 145-161.
 - [4] Cooper, S. and Waldron, S. (2000) *The eigenstructure of the Bernstein operator*. Journal of Approximation Theory, **105**, 133-165.
 - [5] Davis, P. J. (1963) *Interpolation and approximation*. Blaisdell, New York
 - [6] de Boor, C. (1978) *A practical guide to splines*. Applied Mathematical Sciences, 27, Springer-Verlag, New York.
 - [7] Farouki, R. T. (1991) *On the stability of transformations between power and Bernstein polynomials forms*. Computer Aided Geometric Design, **8**, 29-36.
 - [8] Farouki, R.T. (2012) *The Bernstein polynomial basis: a centennial retrospective*. Computer Aided Geometric Design, **29**, 379-419.
 - [9] Impens, C. (2003) *Stirling's series made easy*. Amer. Math'l Monthly, **110**, 730-735.
 - [10] Kato, T. (1995) *Perturbation theory for linear operators*. Springer-Verlag, Berlin, Heidelberg.
 - [11] Laurent, P.-J. (1972) *Approximation et optimisation*. Enseignement des sciences, **13**, Hermann, Paris.
 - [12] Leblanc, A. (2010) *A Bias-reduced approach to density estimation using Bernstein polynomials*. Journal of Nonparametric Statistics, **22**, **4**, 459-475.
 - [13] Leblanc, A. (2012) *On estimating distribution functions using Bernstein polynomials*. Ann. Inst. Stat. Math., **64**, 919-943.
 - [14] Leblanc, A. (2012) *On the boundary properties of Bernstein polynomial estimators of density and distribution functions*. Journal of Statistical Planning and Inference, **142**, 2762-2778.
 - [15] Manté, C. (2012) *Application of iterated Bernstein operators to distribution function and density approximation*. Applied Mathematics and Computation, **218**, 9156-9168.
 - [16] Manté, C. (in revision) *Iterated Bernstein operators for bona fide distribution function and density estimation. Balancing between the iterations number and the polynomial degree*.
 - [17] Sevy, J. C. (1993) *Convergence of iterated boolean sums of simultaneous approximants*. Calcolo, **30**, 41-68.
 - [18] Sevy, J. C. (1995) *Lagrange and least-squares polynomials as limits of linear combinations of iterates of Bernstein and Durrmeyer polynomials*. Journal of Approximation Theory, **80**, 267-271.
 - [19] Vitale, R.A. (1975) *A Bernstein polynomial approach to density function estimation*. Statistical Inference and Related Topics, **2**, 87-99.